Elaboración de distribución de frecuencias y cálculo de estadísticas descriptivas con Gnumeric

Dr. Servio Interiano, noviembre 2020.

Índice

Introducción 2

Gnumeric 2

Aplicación de Gnumeric para el análisis de datos. 4

Preparación de la base de datos. 5

Instrumento de recolección de datos 9

Distribución de frecuencias y cálculo de estadísticas descriptivas empleando Gnumeric. 11

Cuadros para variables cualitativas. 17

Análisis final 20

Introducción

Cuando se inicia con el análisis de un grupo de datos, lo primero que debemos hacer es la presentación de las estadísticas descriptivas, las distribuciones de frecuencias y las gráficas que ayuden a visualizar el comportamiento de los datos.

Es normal encontrar en los cursos, trabajos y/o laboratorios en los que se deba analizar una serie de datos recolectados de una muestra o población.

El procedimiento es básicamente el mismo para todos los grupos.

En el presente documento, vamos a revisar una manera sencilla de hacer esta descripción inicial de los datos aplicando un programa de hoja electrónica multiplataforma y de descarga gratuita, el Gnumeric.

Gnumeric

"Gnumeric es una hoja de cálculo libre que forma parte del entorno de escritorio libre GNOME. La versión 1.0 de Gnumeric fue publicada el 31 de diciembre de 2001. Gnumeric es distribuida como software libre bajo la licencia GNU GPL; Su intención es, junto con todo el software libre, reemplazar al software privativo y hojas de cálculo no libres como Microsoft Excel.

Gnumeric fue creado por Miguel de Icaza, pero él en la actualidad trabaja en otros proyectos. El encargado actual del proyecto es Jody Goldberg." [1]

Características

Gnumeric es capaz de importar y exportar datos en distintos formatos, lo que lo hace compatible con otros programas como Excel, Applix, Quattro Pro, PlanPerfect, Sylk, DIF, Oleo, SC, StarOffice, OpenOffice.org, y Lotus 1-2-3. Su formato nativo es XML, comprimido con gzip. También importa y exporta varios formatos de texto, como tablas HTML o texto separado por comas.

[1] Gnumeric. (2020, 25 de agosto). *Wikipedia, La enciclopedia libre*. Fecha de consulta: 05:13, noviembre 6, 2020 desde https://es.wikipedia.org/w/index.php?title=Gnumeric&oldid=128732825.

Gnumeric se distribuye según las condiciones de la Licencia Pública General de GNU. Ha sido portado a MS Windows (Versiones 2000 y superiores). Para noviembre de 2010 la versión disponible para windows es: 1.12.9. Luego la versión para windows fue discontinuada.

El programa Gnumeric en compañía de Abiword y otros programas es a veces llamado Gnome Office y se presenta como una alternativa ligera a suites de oficina como OpenOffice.org, LibreOffice o KOffice.

El hecho de que sea un programa multiplataforma bajo la Licencia Pública General de GNU, significa que se puede instalar en prácticamente cualquier equipo y utilizarlo sin problemas de licencias o derechos de autor, por lo que es ideal para ser usado en investigación en escuelas y universidades.

Además tiene un potente sistema de análisis estadístico integrado y nos permite hacer distribuciones de frecuencia de manera simple y sencilla.

Lo descargamos directamente desde su página en “gnumeric.org” y lo instalamos siguiendo las instrucciones, dependiendo del sistema operativo que tengamos.

Aplicación de Gnumeric para el análisis de datos.

Vamos a partir de un ejemplo hipotético para ejemplificar de mejor manera cómo hacer el análisis de un par de grupos.

Imaginemos que en un curso X nos dan las instrucciones del cuadro de la derecha.

Finalmente calcularemos el índice de masa corporal (IMC) utilizando la siguiente fórmula:

Peso(kg)/Talla $(m)^2$

La interpretación para el IMC según la OMS:

- <16: Criterio de ingreso a un centro de atención médica
- 16 a 17: Infra peso
- 17 a 18: Bajo peso
- 18 a 25: Peso normal (Saludable)
- 25 a 30: Sobrepeso (Obesidad grado I)
- 30 a 35: Sobrepeso crónico (Obesidad grado II)
- 35 a 40: Obesidad premórbida (Obesidad grado III)
- >40: Obesidad mórbida (Obesidad grado IV)

Se le solicita que, de acuerdo con los parámetros acostumbrados, escriba su informe de laboratorio y realice un análisis exhaustivo de:

- La distribución generada por el peso (hacer curva de distribución normal) y su estadística descriptiva (las dos secciones)
- La proporción de individuos con las distintas variables.
- Establecer si existe diferencias entre hombres y mujeres para el desarrollo de síndrome metabólico a través del IMC y la circunferencia abdominal como predictores de riesgo (puede utilizar una prueba de distribución normal)

Nos piden hacer una investigación acerca de una variable numérica cuantitativa que después deberá presentarse también en categorías. En el caso de nuestro ejemplo el Índice de Masa Corporal, IMC.

Para el ejemplo vamos a suponer que trabajamos con los alumnos de las dos secciones del segundo grado de la carrera de Cirujano Dentista, asignados al curso de Fisiología.

Además de la variable principal que es el IMC, hay otras variables que deben considerarse aunque no se detallen específicamente en las instrucciones de nuestro ejemplo, como lo son la sección (GRUPO), la edad, el género (ya que se solicita que se haga una prueba de diferencia entre hombres y mujeres), el peso y la talla (que son necesarios para la determinación del IMC).

Por lo que deberemos preparar nuestra base de datos de manera que facilite el análisis de los datos.

Preparación de la base de datos.

Antes de poder emplear un programa de análisis en nuestros datos, debemos tener una base de datos adecuada. Vamos a tomar prestado el capítulo 5 del libro "APLICACIÓN PRÁCTICA DE R PARA EL ANÁLISIS DE DATOS EN INVESTIGACIÓN "[2] , en el cual se trata el arreglo de la base de datos para poder usarse en R que es un lenguaje de programación robusto que se emplea para el análisis de datos. En nuestro caso el arreglo general de la base de datos también es aplicable para Gnumeric o cualquier programa de análisis, así que aunque nosotros no vayamos a aplicar R para el análisis de los datos, la estructura y los nombres de las variables es conveniente que lo manejemos de manera correcta.

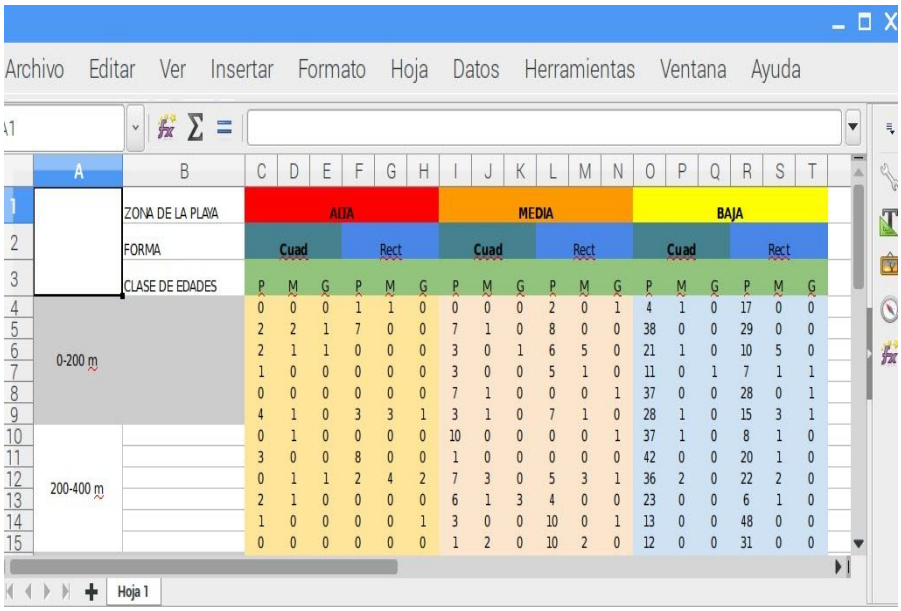
"Uno de los principales problemas que me he encontrado en los 25 años que llevo de enseñar estadística en la universidad, es la estructura de la base de datos que se desea analizar.

En la mayoría de los casos es una hoja de Excel, la cual tiene la peor estructura posible y en algunos casos extremos son tablas elaboradas en Word u otro procesador de texto.

Lo primero que debemos tener en cuenta, es que nuestro archivo de base de datos debe contener el registro de los datos y debe tener cierta estructura para poder analizarlo de manera efectiva.

Nuestro archivo de base de datos debe contener el registro de los datos y debe tener cierta estructura específica para poder analizarlo de manera efectiva.

[2] https://archive.org/details/curso-practico-de-r-v-5 "CURSO PRACTICO DE R V 5" by Servio T. Interiano C.
Publication date 2020-09-02, Usage Attribution-NonCommercial-ShareAlike 4.0 International



	ZONA DE LA PLAYA	ALTA						MEDIA						BAJA					
	FORMA	Cuad			Rect			Cuad			Rect			Cuad			Rect		
	CLASE DE EDADES	P	M	G	P	M	G	P	M	G	P	M	G	P	M	G	P	M	G
0-200 m		0	0	0	1	1	0	0	0	0	2	0	1	4	1	0	17	0	0
		2	2	1	7	0	0	7	1	0	8	0	0	38	0	0	29	0	0
		2	1	1	0	0	0	3	0	1	6	5	0	21	1	0	10	5	0
		1	0	0	0	0	0	3	0	0	5	1	0	11	0	1	7	1	1
		0	0	0	0	0	0	7	1	0	0	0	1	37	0	0	28	0	1
		4	1	0	3	3	1	3	1	0	7	1	0	28	1	0	15	3	1
200-400 m		0	1	0	0	0	0	10	0	0	0	0	1	37	1	0	8	1	0
		3	0	0	8	0	0	1	0	0	0	0	0	42	0	0	20	1	0
		0	1	1	2	4	2	7	3	0	5	3	1	36	2	0	22	2	0
		2	1	0	0	0	0	6	1	3	4	0	0	23	0	0	6	1	0
		1	0	0	0	0	1	3	0	0	10	0	1	13	0	0	48	0	0
		0	0	0	0	0	0	1	2	0	10	2	0	12	0	0	31	0	0

No pueden analizarse datos agrupados o archivos con celdas unidas, textos explicativos, filas o columnas en blanco, etc.

Por lo que revisaremos un poco el cómo preparar de la mejor manera posible, nuestros datos para poder importarlos a R (o cualquier programa de análisis).

Vamos a asumir que tenemos los datos en una hoja electrónica, puede ser esta Excel, Gnumeric, LibreOffice Calc, o cualquier hoja electrónica que se utilice.

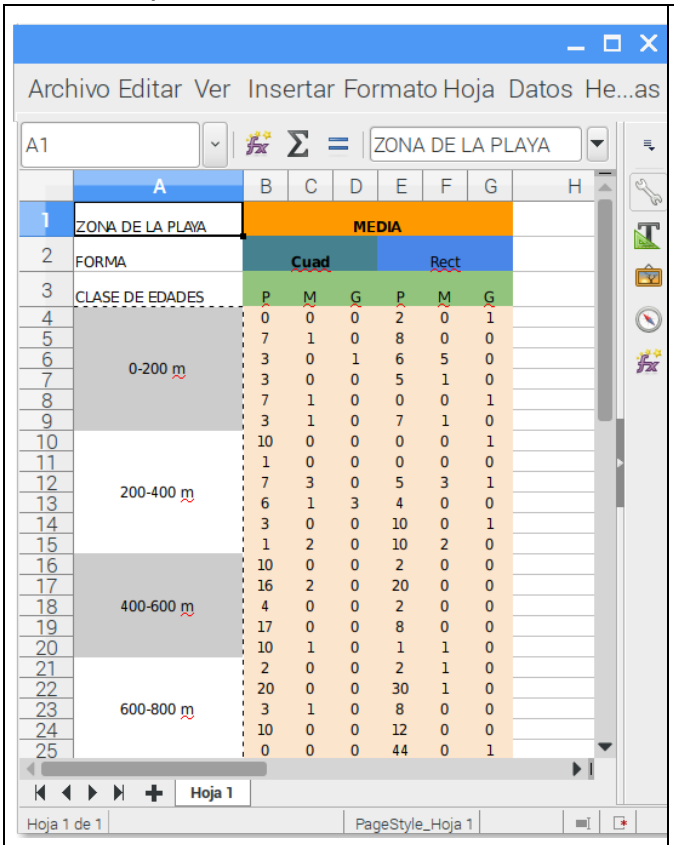
Columnas y Filas

- Las columnas representan a nuestras variables. El nombre de nuestras variables será el encabezado de las columnas y lo vamos a ubicar en la primera fila.
- Nuestras filas representan a nuestras observaciones o unidades de estudio (los sujetos observados). Vamos a emplear la primera columna para ubicar los identificadores de las unidades de estudio. Cada identificador de fila debe ser único, por lo que no debemos tener identificadores repetidos.

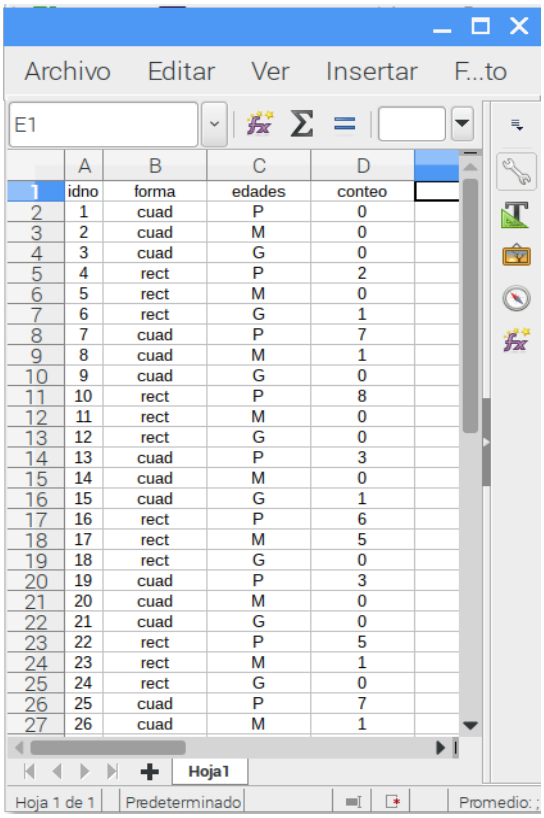
Veamos un ejemplo, tomado de un trabajo de investigación de estudiantes de la Facultad de Biología de la Universidad de San Carlos de Guatemala (pero estoy seguro de que se pueden encontrar ejemplos así en casi cualquier facultad y/o universidad).
En este caso se estaba haciendo un recuento de agujeros de cangrejos en una zona de playa a una altura “Media”, se estaban utilizando dos formas geométricas para el recuento de los agujeros de cangrejo, una cuadrada y otra rectangular y se clasificaron los agujeros dependiendo de su tamaño como pequeño (P), mediano (M) o grande (G) como una estimación de la edad del cangrejo.
En la imagen de la izquierda tenemos la hoja electrónica inicial donde se empezaron a

vaciar los datos.

En la imagen de la derecha tenemos la hoja electrónica ya corregida para poder cargar los datos para su análisis.



Estructura no adecuada



Estructura adecuada

Los nombres de las columnas (nombres de las variables) deben ser compatibles con las reglas para nombrar variables en R.
Normalmente las hojas iniciales de datos tienen problemas con los nombres, que deben ser corregidos antes de proceder a importar los datos a R.

Corrigiendo la base inicial para hacerla compatible con R.

En el encabezado de las columnas (primera fila) se deben evitar nombres con espacios en blanco. Por ejemplo, si estamos analizando resultados de una competencia de atletismo, en vez de escribir “Salto Largo”, podemos escribir “Salto_largo” o “salto.largo”.

Se deben evitar nombres con símbolos especiales: ?, $, *, +, #, (,), -, /, }, {, |, >, <, etc.
Únicamente el guión bajo y el punto pueden ser usados.

Se debe evitar iniciar el nombre de la variable con un número, por ejemplo para el

registro del tiempo en una carrera de 100 metros, debemos evitar “100m” y podemos usar “Tiempo100m” o “T100m”.

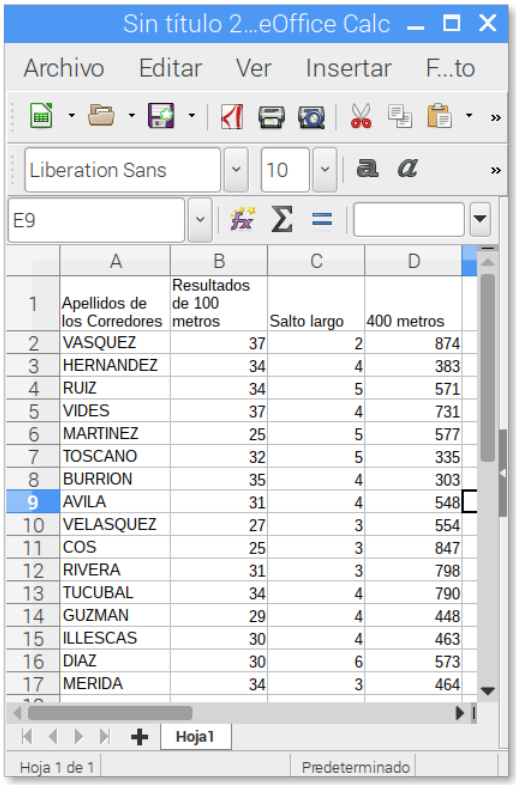
Los nombres de las columnas deben ser únicos, no se permiten nombres duplicados.
Es conveniente no poner nombres muy largos, porque complican la sintaxis en el momento del análisis (Generalmente estos nombres corresponden con el nombre de la variable).

Hay que tener presente que:

- R es sensible a las minúsculas y mayúsculas, lo que implica que “tiempo” es distinto a “Tiempo”.
- Se debe evitar dejar filas en blanco.
- Se debe borrar cualquier comentario.
- Se debe reemplazar los valores faltantes por NA (Not Available).
- Para las fechas se debe usar el formato de cuatro dígitos en los años, así se debe usar “01/01/2016” en vez de “01/01/16”.

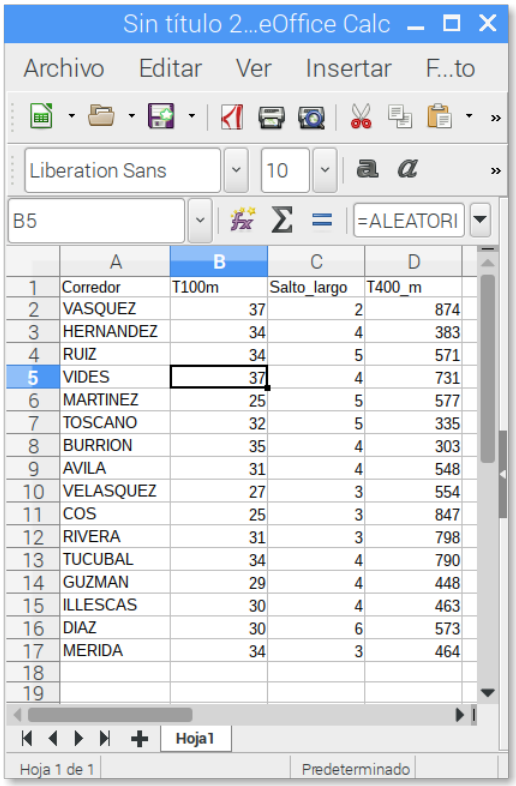
En las siguientes figuras vemos un ejemplo de nombres incorrectos y nombres correctos para importar los datos a R.

En la figura de la izquierda tenemos nombres de variables incorrectos, mientras que en la de la derecha se muestran nombres compatibles con R para importar los datos.



	A	B	C	D
1	Apellidos de los Corredores	Resultados de 100 metros	Salto largo	400 metros
2	VASQUEZ	37	2	874
3	HERNANDEZ	34	4	383
4	RUIZ	34	5	571
5	VIDES	37	4	731
6	MARTINEZ	25	5	577
7	TOSCANO	32	5	335
8	BURRION	35	4	303
9	AVILA	31	4	548
10	VELASQUEZ	27	3	554
11	COS	25	3	847
12	RIVERA	31	3	798
13	TUCUBAL	34	4	790
14	GUZMAN	29	4	448
15	ILLESCAS	30	4	463
16	DIAZ	30	6	573
17	MERIDA	34	3	464

Nombres de las variables incorrectos



	A	B	C	D
1	Corredor	T100m	Salto_largo	T400_m
2	VASQUEZ	37	2	874
3	HERNANDEZ	34	4	383
4	RUIZ	34	5	571
5	VIDES	37	4	731
6	MARTINEZ	25	5	577
7	TOSCANO	32	5	335
8	BURRION	35	4	303
9	AVILA	31	4	548
10	VELASQUEZ	27	3	554
11	COS	25	3	847
12	RIVERA	31	3	798
13	TUCUBAL	34	4	790
14	GUZMAN	29	4	448
15	ILLESCAS	30	4	463
16	DIAZ	30	6	573
17	MERIDA	34	3	464
18				
19				

Nombres de las variables correctos

Instrumento de recolección de datos

Otro aspecto que es importante revisar es el instrumento de recolección de datos a utilizar.

Generalmente las fichas que se emplean son como el ejemplo siguiente:

LABORATORIO No. 4

Nombre	Talla	Peso	Cintura

El nombre del sujeto o unidad de estudio no es un dato que nos interese a menos que se piense hacer una rectificación de datos, por lo que podría utilizarse un número de identificación como el carné o un IDNO .

En nuestro ejemplo es importante determinar el GRUPO o sección al que pertenece nuestro sujeto, pero no está la columna en la ficha, por lo que la ficha realmente no nos va a ser de mucha utilidad.

Falta también los datos del género y la altura.

Así mismo se recolectan los datos de talla, peso y cintura. Pero en la investigación interesa el IMC y la medida de la cintura, por lo que podría anotarse el valor del IMC en vez de la talla y el peso.

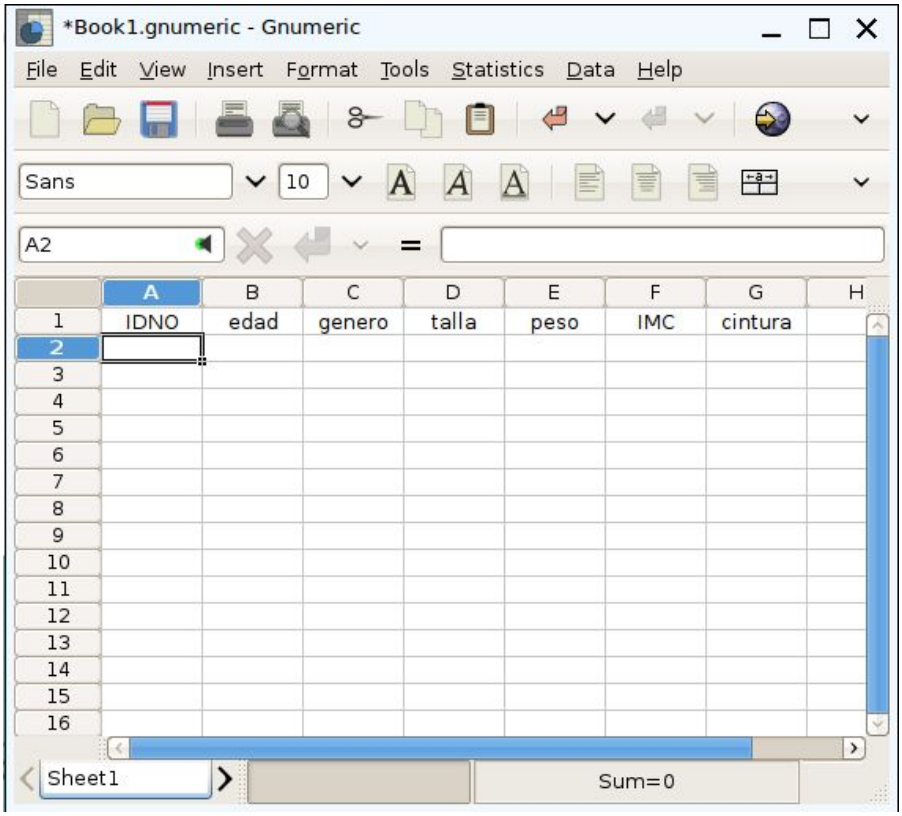
Si por cuestiones de tiempo en el momento del trabajo de campo no se puede calcular, entonces se toman los datos de talla y peso, dejando el espacio para anotar el IMC.

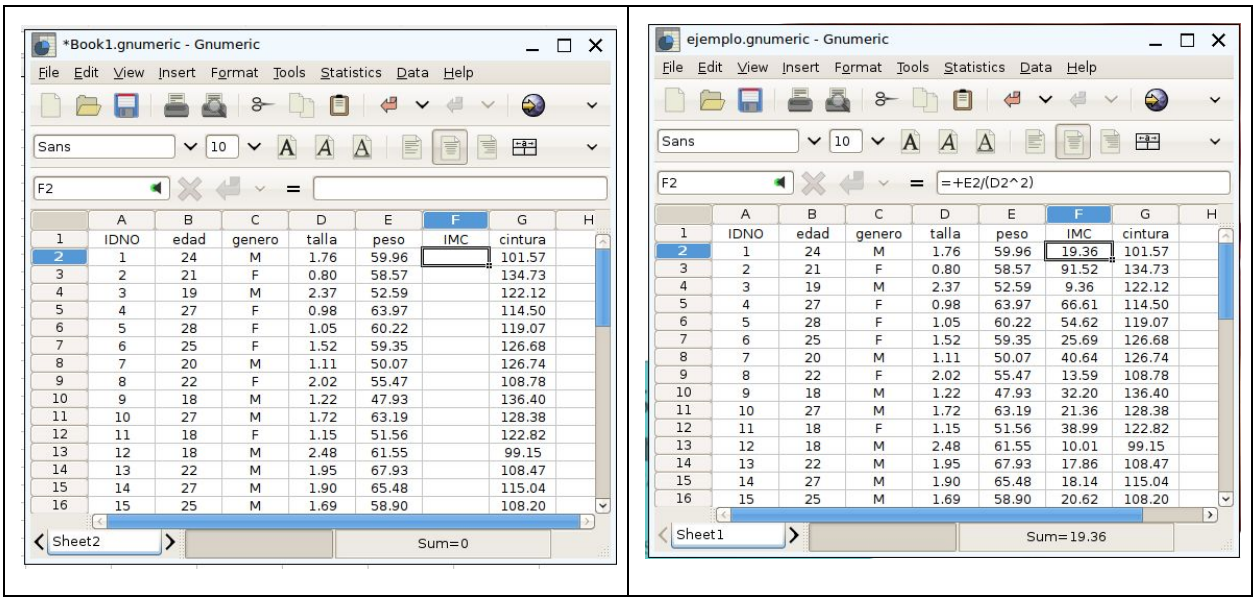
Un ejemplo de como debería ser el instrumento de recolección de datos es el siguiente:

Nombre de quien anota:_____ Fecha: _____

IDNO	Edad	Género	Talla	Peso	IMC	Cintura

Los datos pueden registrarse directamente en la hoja de cálculo, que es a donde finalmente se transfieren para elaborar la base de datos.





IDNO	edad	genero	talla	peso	IMC	cintura
1	24	M	1.76	59.96		101.57
2	21	F	0.80	58.57		134.73
3	19	M	2.37	52.59		122.12
4	27	F	0.98	63.97		114.50
5	28	F	1.05	60.22		119.07
6	25	F	1.52	59.35		126.68
7	20	M	1.11	50.07		126.74
8	22	F	2.02	55.47		108.78
9	18	M	1.22	47.93		136.40
10	27	M	1.72	63.19		128.38
11	18	F	1.15	51.56		122.82
12	18	M	2.48	61.55		99.15
13	22	M	1.95	67.93		108.47
14	27	M	1.90	65.48		115.04
15	25	M	1.69	58.90		108.20

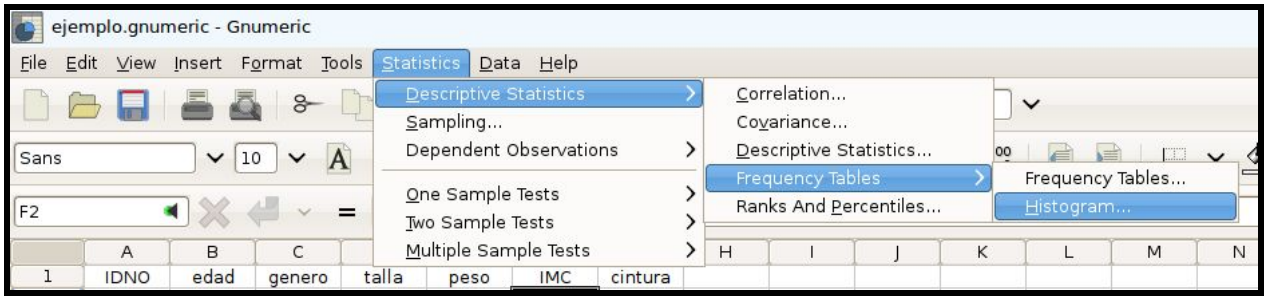
IDNO	edad	genero	talla	peso	IMC	cintura
1	24	M	1.76	59.96	19.36	101.57
2	21	F	0.80	58.57	91.52	134.73
3	19	M	2.37	52.59	9.36	122.12
4	27	F	0.98	63.97	66.61	114.50
5	28	F	1.05	60.22	54.62	119.07
6	25	F	1.52	59.35	25.69	126.68
7	20	M	1.11	50.07	40.64	126.74
8	22	F	2.02	55.47	13.59	108.78
9	18	M	1.22	47.93	32.20	136.40
10	27	M	1.72	63.19	21.36	128.38
11	18	F	1.15	51.56	38.99	122.82
12	18	M	2.48	61.55	10.01	99.15
13	22	M	1.95	67.93	17.86	108.47
14	27	M	1.90	65.48	18.14	115.04
15	25	M	1.69	58.90	20.62	108.20

Cargamos los datos y hacemos que el programa calcule el IMC, directamente en la hoja de cálculo.

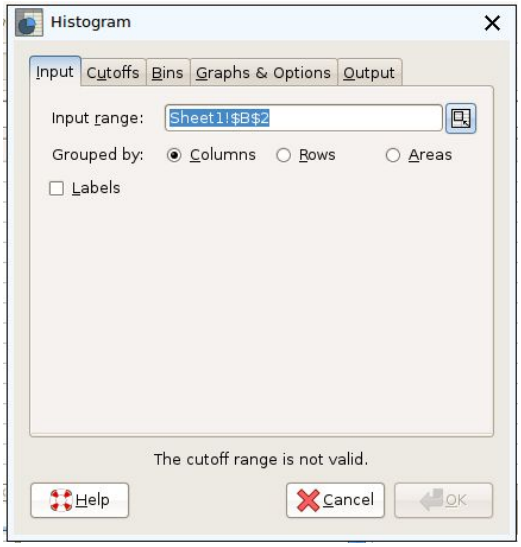
Y ya estamos listos para empezar el análisis.

Distribución de frecuencias y cálculo de estadísticas descriptivas empleando Gnumeric.

Aquí es cuando la ventaja del programa se hace visible. Para elaborar una distribución de frecuencias de la variable “edad”, por ejemplo, utilizamos la función



Statistics Descriptive Statistics Frequency Tables Histogram...	Este comando nos permite fabricar una distribución de frecuencias para variables cuantitativas, con solo indicar el rango de datos, el número de intervalos, el dato menor y el dato mayor.



La primera pestaña de la ventana que se abre, nos permite seleccionar el rango de datos, así como indicar si nuestros datos están en columnas, filas o áreas combinadas de columnas y filas.

Como vimos en la sección de preparación de la base de datos, es recomendable que estén en columnas.

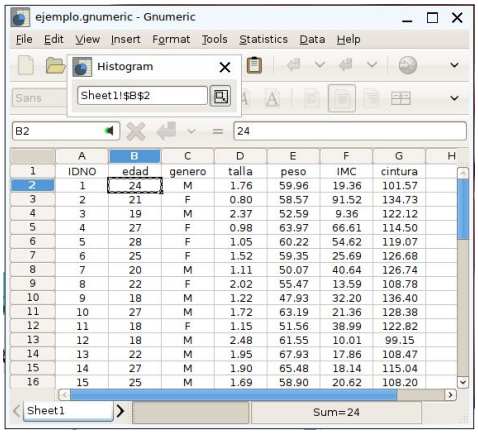
Tambíen nos permite indicarle al programa si la primera fila es de etiqueta, esto marcando el cuadrado donde dice “etiquetas” [Labels]

Para seleccionar el rango de datos hacemos click en el icono a la derecha donde dice “Input range”

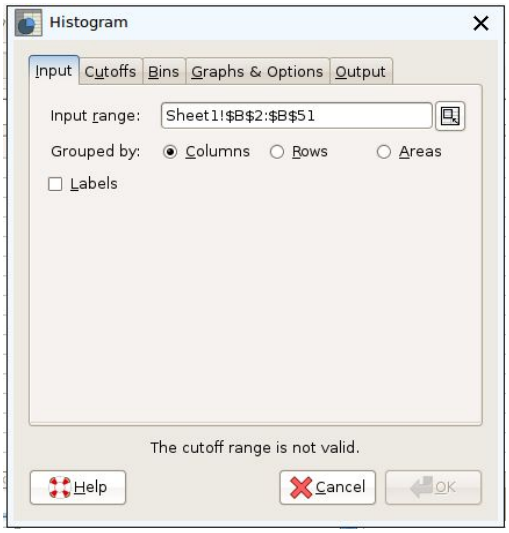
Luego nos ubicamos en el inicio de la columna que vamos a trabajar, en este ejemplo la variable edad.

Y seleccionamos la columna completa.

Al finalizar la selección, hacemos click en el icono a la derecha de donde se marca la selección y regresamos a la ventana donde estábamos anteriormente.



	IDNO	edad	genero	talla	peso	IMC	cintura
1	1	24	M	1.76	59.96	19.36	101.57
2	2	21	F	0.80	58.57	91.52	134.73
3	3	19	M	2.37	52.59	9.36	122.12
4	4	27	F	0.98	63.97	66.61	114.50
5	5	28	F	1.05	60.22	54.62	119.07
6	6	25	F	1.52	59.35	25.69	126.68
7	7	20	M	1.11	50.07	40.64	126.74
8	8	22	F	2.02	55.47	13.59	108.78
9	9	18	M	1.22	47.93	32.20	136.40
10	10	27	M	1.72	63.19	21.36	128.38
11	11	18	F	1.15	51.56	38.99	122.82
12	12	18	M	2.48	61.55	10.01	99.15
13	13	22	M	1.95	67.93	17.86	108.47
14	14	27	M	1.90	65.48	18.14	115.04
15	15	25	M	1.69	58.90	20.62	108.20



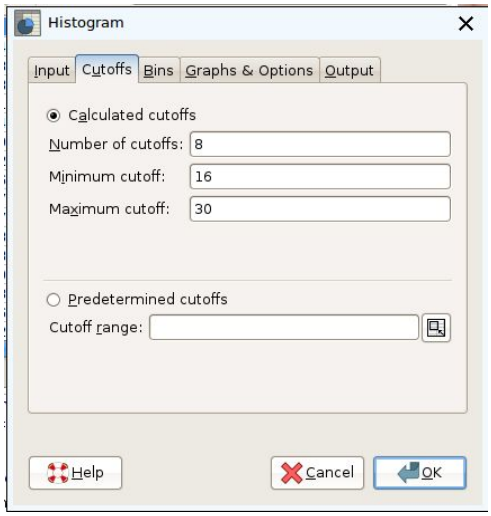
Ya con nuestros datos seleccionados, nos desplazamos a la siguiente pestaña: “Cutoffs”, que es donde le vamos a indicar al programa el número de intervalos y los límites inferior y superior.

La primera casilla [Number of cutoffs] es para el número de intervalos que vamos a querer.

La segunda [Minimum cutoff] es el límite inferior, que generalmente es nuestra observación más pequeña.

La tercera casilla [Maximum cutoff] es el límite superior o dato máximo que vamos a incluir en la distribución de frecuencias.

Para nuestro ejemplo, deseamos 8 intervalos, el dato menor es 16 y el dato mayor es 30, así que seleccionamos esos valores.



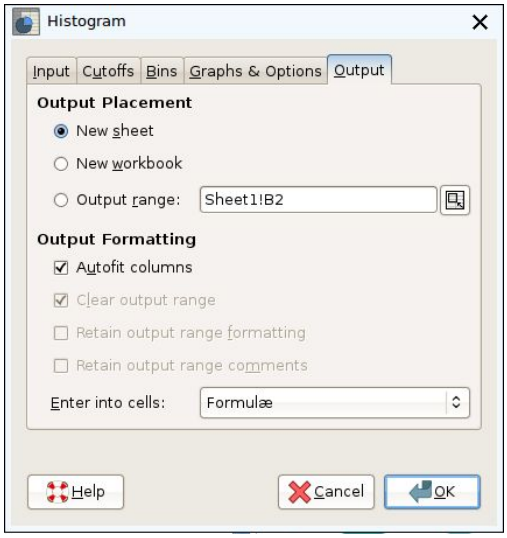
La siguiente pestaña nos permite definir la forma que queremos que se comporten nuestros intervalos, ya sea dejando abiertos los extremos inferior y superior, dejando abierto el extremo superior y exacto o aproximado el extremo inferior, dejando abierto el extremo inferior y exacto o aproximado el extremo superior, etc.

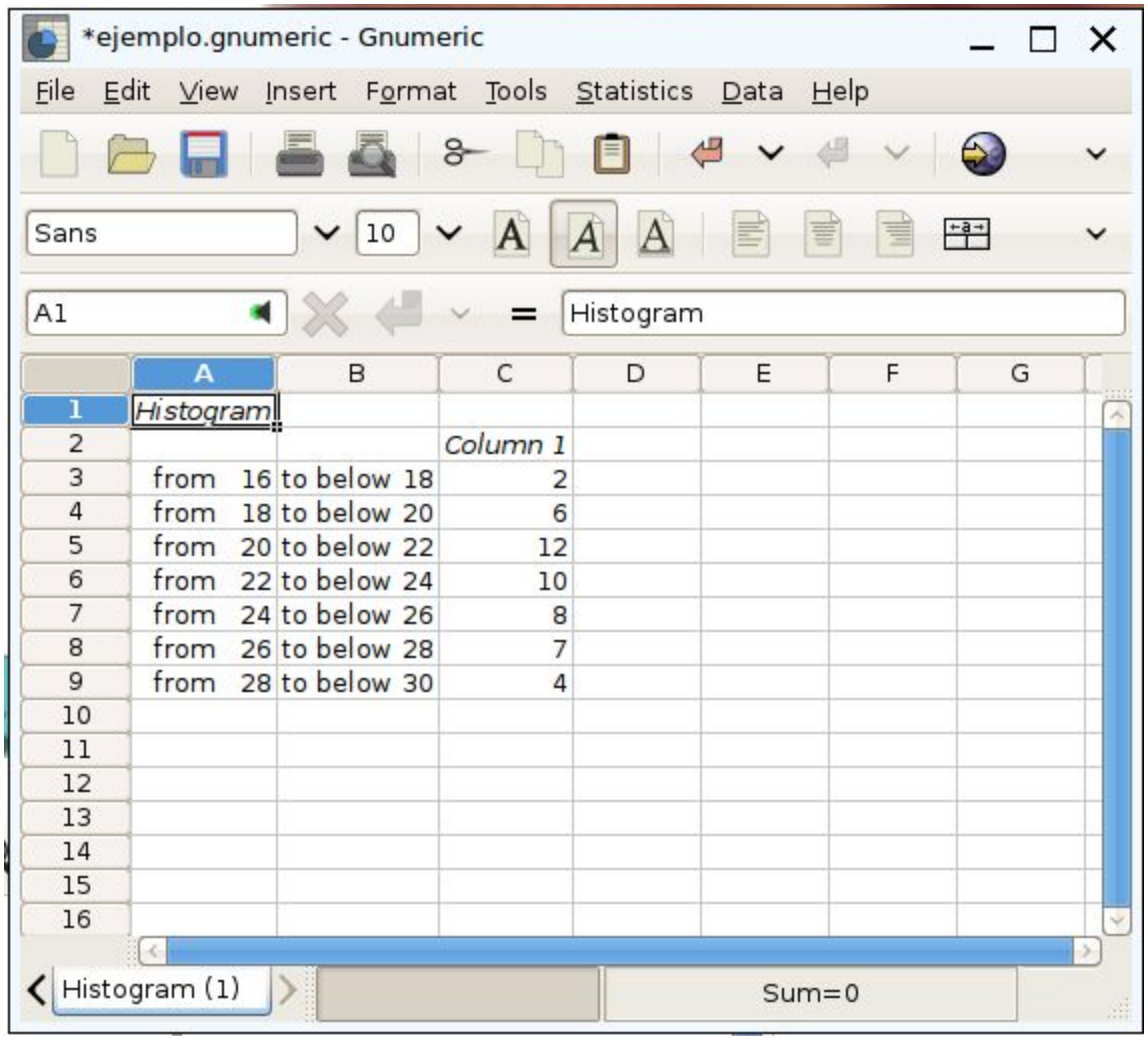
En nuestro caso, seleccionamos no dejar abierto ninguno de los extremos.

Y por último definimos donde queremos la salida de nuestra distribución de frecuencias.

La opción por defecto es en una nueva hoja [New sheet] y es una buena opción, por lo que sugeriría no modificar nada en esta ventana.

Le damos click en [OK] y listo!





El programa nos abre una nueva hoja electrónica con nuestra distribución de frecuencias en donde nos indica que:

desde 16 hasta debajo de 18 tenemos 2 datos.
desde 18 hasta debajo de 20 tenemos 6 datos.
desde 20 hasta debajo de 22 tenemos 12 datos. Etc.

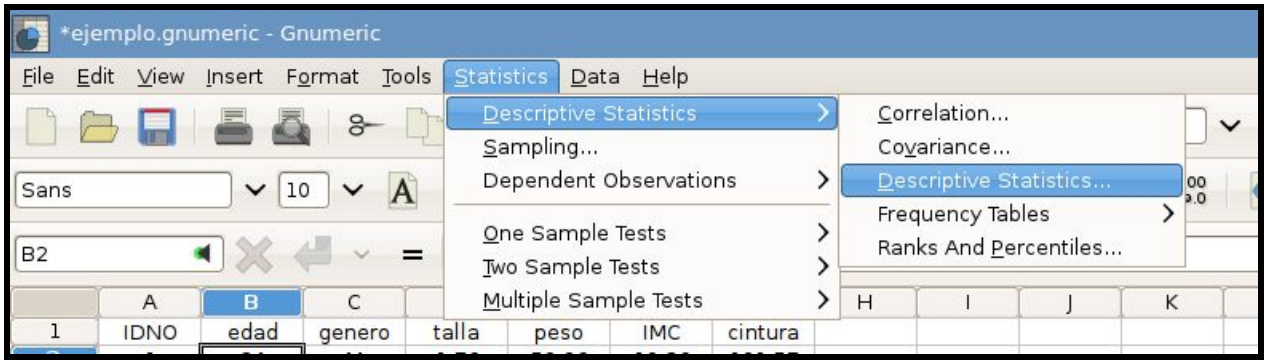
Lo que podemos convertir fácilmente en una distribución de frecuencias, como se muestra en el cuadro de la derecha.

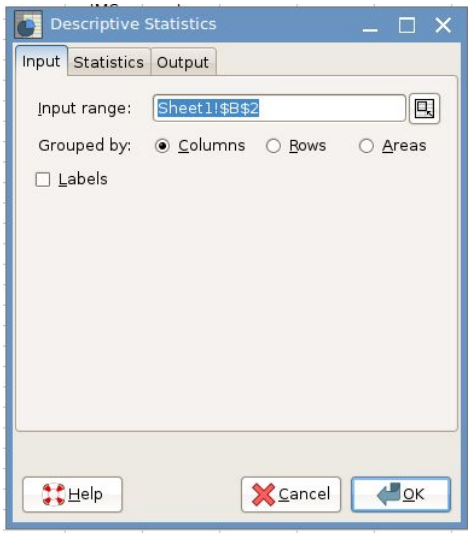
Edades			f	fr	fa
16	-	17	2	0.04	2
18	-	19	6	0.12	8
20	-	21	12	0.24	20
22	-	23	10	0.20	30
24	-	25	8	0.16	38
26	-	27	7	0.14	45
28	-	30	5	0.10	50
			50		

Pero eso no es todo.

De una manera igualmente sencilla podemos calcular todas las estadísticas descriptivas.

Seleccionamos: [Statistics] , [Descriptive Statistics], [Descriptive Statistics…]





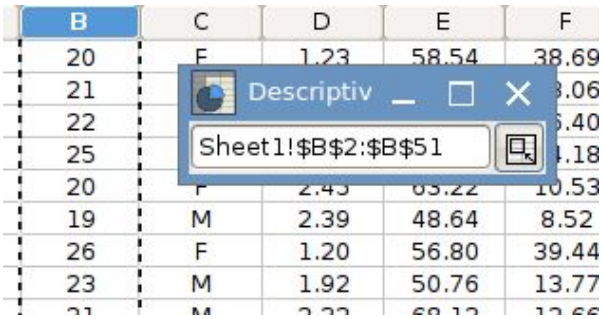
Esto nos abre una ventana de diálogo como en la sección anterior, pero mucho más sencilla.

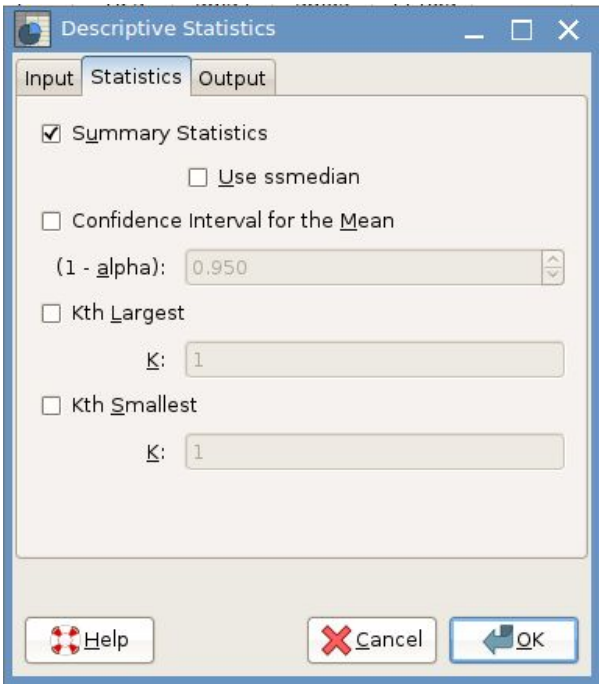
Seleccionamos el rango de datos para analizar a través del icono a la derecha de [Input range].

Los otros botones y opciones son iguales a los descritos anteriormente.

Marcamos la columna que nos interesa, en este caso la columna de la edad, desde el primer dato hasta el último y finalizamos la selección haciendo click en el icono a la derecha de la ventana de selección.

Esto nos regresa a la ventana anterior y seleccionamos la siguiente pestaña.

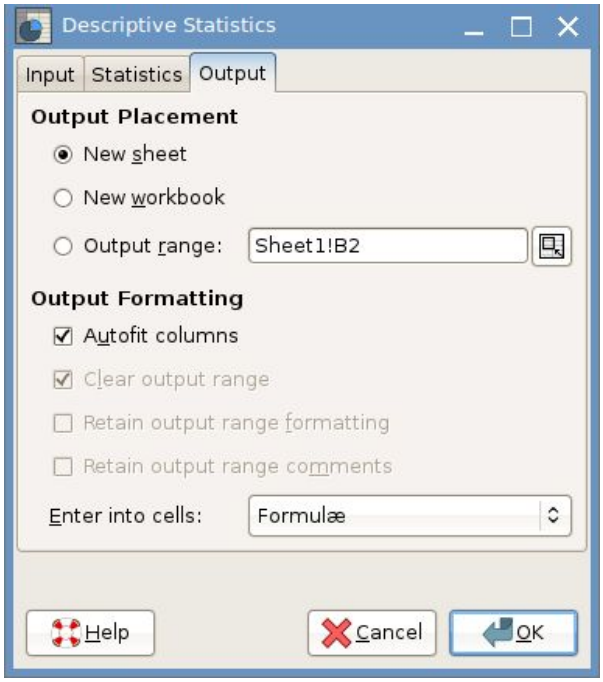




Aquí la opción por defecto es [Summary Statistics] , así que no debemos hacer nada más que darle [OK]

La última pestaña tampoco necesita ningún ajuste, y aquí la muestro solo para fines de descripción.

La opción de que la salida de los datos sea en una nueva hoja es la más adecuada, así que solo le damos [OK]



	A	B
1		Column 1
2	Mean	22.78
3	Standard Error	0.47491352596326
4	Median	22.5
5	Mode	21
6	Standard Deviation	3.35814574685835
7	Sample Variance	11.2771428571429
8	Kurtosis	−0.74756243861829
9	Skewness	0.11127941583941
10	Range	14
11	Minimum	16
12	Maximum	30
13	Sum	1139
14	Count	50
15		

El resultado que nos aparece en la nueva hoja, es un cálculo de las principales estadísticas descriptivas:
La media aritmetica, el error estándar, la mediana, la moda, la desviación estándar de una muestra, la varianza para una muestra, la curtósis, el sesgo, el rango de los datos, el mínimo, el máximo, la sumatoria de los datos y el número de datos.

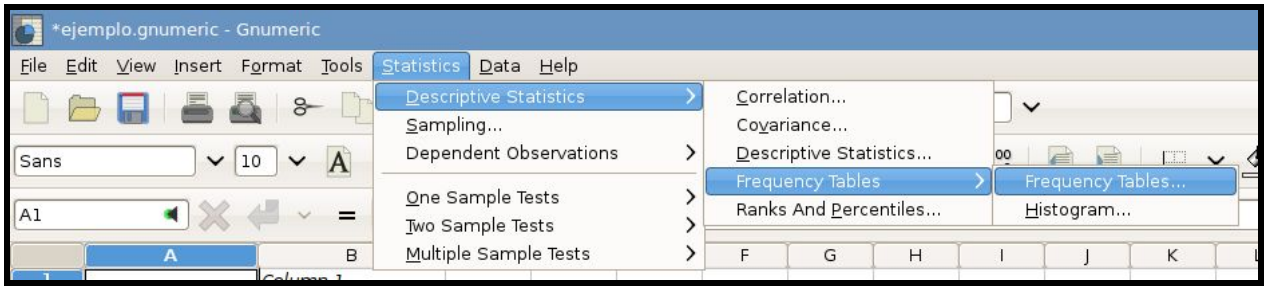
Si se está trabajando con una población, transformamos la varianza, multiplicandola por “n-1” y dividiéndola entre la “n”. A esta varianza corregida, le sacamos la raíz cuadrada y con esto tenemos la desviación estándar de la población.

Cuadros para variables cualitativas.

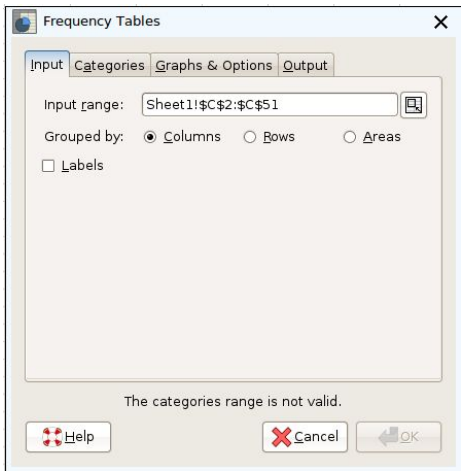
La elaboración de cuadros para variables cualitativas es más sencilla que una distribución de frecuencias y basta con contar cada categoría.

Gnumeric también puede elaborar este tipo de cuadros, a través de seleccionar los datos.

Para hacer esto utilizamos [Statistics], [Descriptive Statistics], [Frequency Tables], [Frequency Tables…]



Esto nos abre una ventana de diálogo, donde podemos seleccionar nuestros datos cualitativos y ordenarlos en un cuadro de frecuencias.

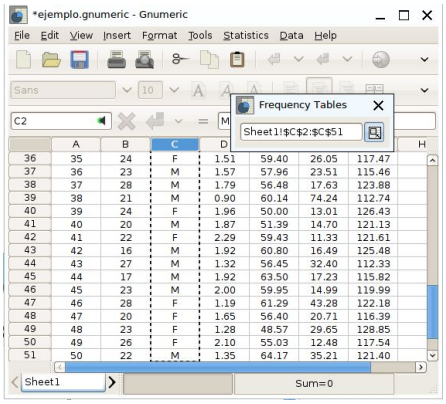


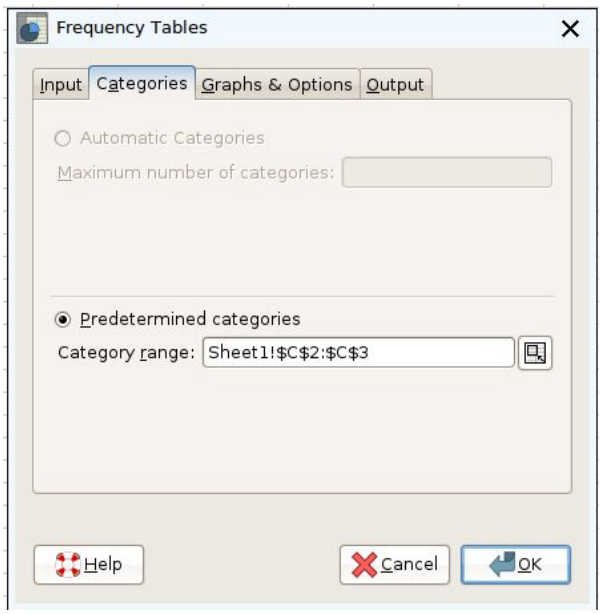
Seleccionamos el rango de datos, de manera similar a como hemos hecho anteriormente.

Lo único que varía es que los datos que vamos a seleccionar ahora no son números, sino que categorías.

Seleccionamos la columna de la variable género, donde los valores son M para masculino y F para femenino.

Al finalizar de seleccionar nuestra columna, hacemos click en el icono a la derecha de la ventana de selección.





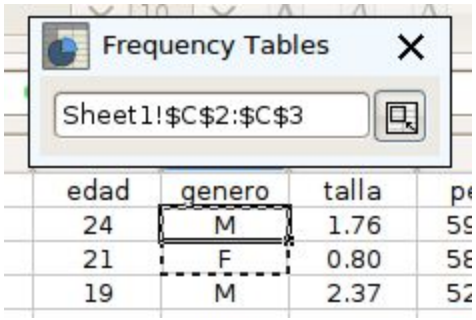
En la siguiente pestaña, en [Categories] , seleccionamos las categorías que deseamos para presentar en nuestro cuadro.

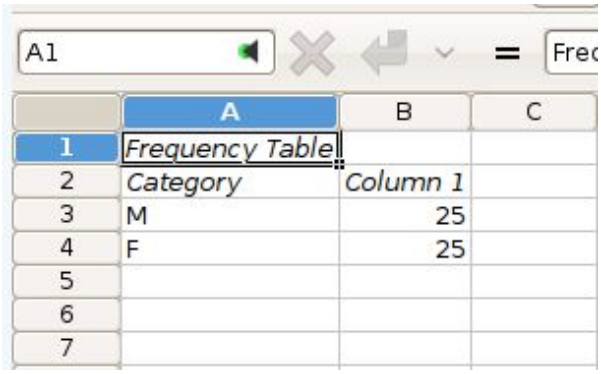
Esto podemos hacerlo escribiendo las categorías en la hoja electrónica, o seleccionando directamente las categorías de nuestros datos.

Como en este caso el género solo tiene dos categorías, lo haremos directamente desd los datos.

Abrimos la ventana de selección, haciendo click en el icono a la derecha de [Category range]

Seleccionamos M y F como nuestras categorías en el cuadro y finalizamos la selección haciendo click en el icono a la derecha de la ventana de selección.





El resultado es un recuento de cada categoría, en nuestro caso M (masculino) tiene una frecuencia de 25 y F (femenino) una frecuencia de 25.

El procedimiento es el mismo para cualquier variable categórica.

Análisis final

Aplicando lo revisado con el gnumeric la elaboración de distribuciones de frecuencias y el análisis de datos es muy sencillo.

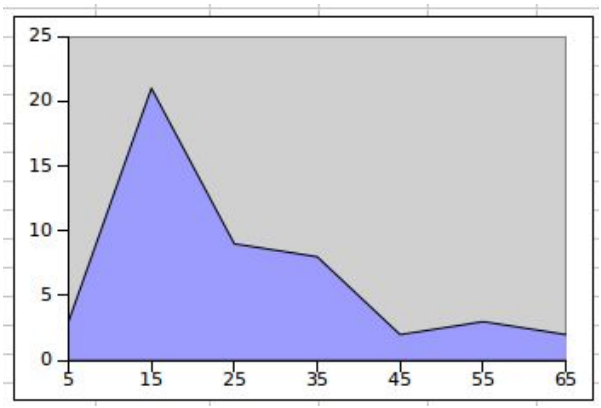
En nuestro ejemplo, donde nos interesa el IMC:

	A	B	C
1	*Histogram*		
2			*Column 1*
3	from 0	to below 10	3
4	from 10	to below 20	21
5	from 20	to below 30	9
6	from 30	to below 40	8
7	from 40	to below 50	2
8	from 50	to below 60	3
9	from 60	to below 70	2
10			

x	f
5	3
15	21
25	9
35	8
45	2
55	3
65	2

Calculamos el cuadro de frecuencia para elaborar nuestra distribución de frecuencias.

Con la distribución de frecuencias elaborada, calculamos la marca de clase de cada intervalo y utilizamos estos datos junto a la frecuencia para elaborar un polígono de frecuencias.



	A	B	C
1		*IMC*	
2	*Mean*	28.1884937198637	
3	*Standard Error*	2.67733478678483	
4	*Median*	20.6678707832822	
5	*Mode*	#N/A	
6	*Standard Deviation*	18.9316158324219	
7	*Sample Variance*	358.406078026409	
8	*Kurtosis*	1.93727429673164	
9	*Skewness*	1.48187374638959	
10	*Range*	83.6653322261204	
11	*Minimum*	7.85555527387958	
12	*Maximum*	91.5208875	
13	*Sum*	1409.42468599319	
14	*Count*	50	
15			

El polígono de frecuencias nos muestra un comportamiento asimétrico positivo.

Las estadísticas descriptivas nos brindan el sesgo y la curtosis para verificar el comportamiento del grupo.
En este caso el sesgo= 1.48 y sus límites +/- 0.69
por lo que es asimétrico positivo.
La curtosis es 1.94 y sus límites son +/- 1.38, por lo que el grupo es leptocúrtico.
El grupo no es normal.